



SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Neuhold, Lisa
Killar, Loran

(ii) TITLE OF THE INVENTION: TRANSGENIC ANIMAL MODEL FOR
DEGENERATIVE DISEASES OF CARTILAGE

(iii) NUMBER OF SEQUENCES: 21

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Darby & Darby PC
- (B) STREET: 805 Third Avenue
- (C) CITY: New York
- (D) STATE: NY
- (E) COUNTRY: USA
- (F) ZIP: 10022

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: DOS
- (D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: 08/994,689
- (B) FILING DATE: 1997-12-19
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Green, Reza
- (B) REGISTRATION NUMBER: 38,475
- (C) REFERENCE/DOCKET NUMBER: 0630/0D532

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 212-527-7700
- (B) TELEFAX: 212-753-6237
- (C) TELEX: 236687

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 471 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met His Pro Gly Val Leu Ala Ala Phe Leu Phe Leu Ser Trp Thr His
1 5 10 15
Cys Arg Ala Leu Pro Leu Pro Ser Gly Gly Asp Glu Asp Asp Leu Ser
20 25 30
Glu Glu Asp Leu Gln Phe Ala Glu Arg Tyr Leu Arg Ser Tyr Tyr His
35 40 45
Pro Thr Asn Leu Ala Gly Ile Leu Lys Glu Asn Ala Ala Ser Ser Met
50 55 60
Thr Glu Arg Leu Arg Glu Met Gln Ser Phe Phe Gly Leu Glu Val Thr
65 70 75 80
Gly Lys Leu Asp Asp Asn Thr Leu Asp Val Met Lys Lys Pro Arg Cys
85 90 95
Gly Val Val Asp Val Gly Glu Tyr Asn Val Phe Pro Arg Thr Leu Lys
100 105 110
Trp Ser Lys Met Asn Leu Thr Tyr Arg Ile Val Asn Tyr Thr Pro Asp
115 120 125
Met Thr His Ser Glu Val Glu Lys Ala Phe Lys Lys Ala Phe Lys Val
130 135 140
Trp Ser Asp Val Thr Pro Leu Asn Phe Thr Arg Leu His Asp Gly Ile
145 150 155 160
Ala Asp Ile Met Ile Ser Phe Gly Ile Lys Glu His Gly Asp Phe Tyr
165 170 175
Pro Phe Asp Gly Pro Ser Gly Leu Leu Ala His Ala Phe Pro Pro Gly
180 185 190
Pro Asn Tyr Gly Gly Asp Ala His Phe Asp Asp Asp Glu Thr Trp Thr
195 200 205
Ser Ser Ser Lys Gly Tyr Asn Leu Phe Leu Val Ala Ala His Glu Phe
210 215 220
Gly His Ser Leu Gly Leu Asp His Ser Lys Asp Pro Gly Ala Leu Met
225 230 235 240
Phe Pro Ile Tyr Thr Tyr Thr Gly Lys Ser His Phe Met Leu Pro Asp
245 250 255
Asp Asp Val Gln Gly Ile Gln Ser Leu Tyr Gly Pro Gly Asp Glu Asp

260	265	270
Pro	Asn	Pro
Lys	His	Pro
Pro	Lys	Thr
Asp	Lys	Cys
275	280	285
Ser	Leu	Asp
Ala	Ile	Thr
Ser	Leu	Arg
Gly	Glu	Thr
290	295	300
Asp	Arg	Phe
Phe	Trp	Arg
305	310	315
Leu	His	Pro
Gln	Gln	Val
Asp	Ala	Glu
320	325	330
Phe	Leu	Thr
Lys	Ser	Phe
Trp	Pro	Glu
335	340	345
Leu	Pro	Asn
Arg	Ile	Asp
350	355	360
Ala	Tyr	Glu
His	Pro	Ser
His	Asp	Leu
Ile	Phe	Ile
Phe	Arg	Gly
365	370	375
Gly	Tyr	Asp
380	385	390
Trp	Ala	Leu
Asn	Gly	Tyr
395	400	405
Asp	Ile	Leu
Thr	Gly	Glu
410	415	420
Lys	Thr	Leu
Leu	Leu	Phe
425	430	435
Pro	Ser	Ser
Gly	Ile	Ile
Asp	Phe	Gly
445	450	455
Val	Asp	Asp
460	465	470
Asn	Ser	Ile
Ile	Leu	Trp
Cys		

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CTCGAGTTA	CCACTCCCTA	TCAGTGATAG	AGAAAAGTGA	AAGTCGAGTT	TACCACTCCC	60
TATCAGTGAT	AGAGAAAAGT	GAAAGTCGAG	TTTACCACTC	CCTATCAGTG	ATAGAGAAAA	120
GTGAAAGTCG	AGTTTACAC	TCCCTATCAG	TGATAGAGAA	AAGTGAAAGT	CGAGTTTACC	180
ACTCCCTATC	AGTGATAGAG	AAAAGTGAAA	GTCGAGTTA	CCACTCCCTA	TCAGTGATAG	240
AGAAAAGTGA	AAGTCGAGTT	TACCACTCCC	TATCAGTGAT	AGAGAAAAGT	GAAAGTCGAG	300
CTCGGTACCC	GGGTCGAGTA	GGCGTGTACG	GTGGGAGGCC	TATATAAGCA	GAGCTCGTT	360
AGTGAACCGT	CAGATCGCCT	GGAGACGCCA	TCCACGCTGT	TTTGACCTCC	ATAGAAGACA	420

CCGGGACCGA TCCAGCCTCC GCGGCCCGA ATTAGCTTGA TATCGAATT

470

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3479 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGTACCACTA	GTAAGCTTAG	ATCCACTGTC	TGGGATTATA	TCAGGACAAC	CGAACGCCTGG	60
AAAGTGTATT	AGGTAGAGCA	TTTCTTCCA	CGTGTGGG	CACGTTCCG	ACAGCTAGGA	120
TTCCAGCTCT	GTCTTGTAT	GTTACAGACT	GTAAATCAAT	CGCAGGTGAA	ACTGTTGGA	180
CAGTAGGTGG	GGATCAAAGA	CCCTCCGCC	GTGAGACTCT	AGGCCTTTC	CCCTGCCACC	240
AGCCTGTCTC	CAGAGATGCT	CTGGAAGGAG	GCAGGGCCCGG	GCAGGTCTTC	TGCTCTTAG	300
CGTGGCGGAC	GCAGGCGGCGG	GGCAGGGCT	GGAGCAGAGA	GCGCTGCAGT	GATAGAACTT	360
TCTGACCCCG	CTGCGCAGGG	CGGCAGGGTG	GCAGGGTGGC	AGGGTGGCGA	GCTAAGCCAG	420
AGCCGAACGC	TGGAGCTCTG	GGAGGAACAT	CGAAGGTTTG	TATGTGGTCT	GAGATCGGCC	480
TGACTATATT	TTTTTGTCTT	AAATTTGCAA	GCACACACCC	ACAAAGCTGC	GGTCTTGACC	540
GGTATTCTTT	ATAGAGCGCA	ATGGAGTGAG	CTGAGTGTCT	AAACGATTTC	CCTAATTCTAT	600
CTGATAGCAG	AGGCGCTCTC	CTAATTGGCG	AAGAGCTGCC	TCATGTCCGC	AACTTTTGG	660
CAGAGTGAAT	TCCACAGCTT	TGTGTGTGTG	TGTGGGGGGG	GGTGTAAAGG	GTGTCTAAAA	720
CTTCGGTCT	CCTACTATTC	TGTATCTCGA	CCGGTTGGTT	TTACACCCCG	GCTCATCTCA	780
TCAACGCAA	CACCCCCACT	CTCCTATGGA	CCCAAGGACC	TGACGTGGGG	GAAGGGGAC	840
ATTAGGAATG	TCAGAAACCT	AGAGTCCACG	CTCCTCCTCT	CCATCTTCC	ACGAGTTGG	900
GAAACTTCTT	GGCTGCGAAG	ACTTGACCC	ACATCTGCAT	TTCTCAGCCC	CAGCTCCAA	960
AAGTGCTGCA	GGTCGGGAG	GGGAGACCTC	AGTCCTCCTT	TGTGAGGCTT	GTTCGCGTTG	1020
AGGGATTGGC	AGCGATGGCT	TCCAGATGGG	CTGAAACCCCT	GCCCGTATT	ATTTAAACTG	1080
GTTCCTCGTG	GAGAGCTGTG	AATCGGGCTC	TGTATGCGCT	CGAGAAAAGC	CCCATTCTATG	1140
AGAGGCAAGG	CCCAGTGGGT	CCCCCGACT	CCCCGACCC	CCTCTCCCAC	AATATATCCC	1200
CCCTCCCTGT	GCCCGCCTGC	CGCCACCTCC	CGGGCTCCGG	CCCCGCGCGC	AGCGGCGACG	1260
AAGCAACACA	GTTCCCCGAA	AGAGGTAGCT	TTTAATTGG	CCAGCCACAA	AGAATCACTT	1320
ATGCCGCACG	GCGGTAACGA	GGGGAACCGG	ATCGGGCGGC	CAGGATGCTA	TCTGTGTAGC	1380
CCTTTCTGT	CCACAATTAG	GGTGGTGCTG	GCTTCCTCCG	ACCGCACCTA	GGCGATCTGG	1440
TTACACTGTT	GGCTCCTTTC	TTGGGCAGTC	ATTTAACCT	ACTTTTACT	CTACGAATGT	1500
CTGTCTGATG	GAGGGCTGTG	TCCGGAGCCC	CATCCACAAA	GAGTCAGCCA	GCAGCTCTCA	1560
CACCCGGCTG	GATCTCATAT	GGTGCACTCT	CAGTACAATC	TGCTCTGATG	CCGCATAGTT	1620
AAGCCAGCCA	AGCTAGCTT	CGCAAGCTAG	CTTGCAGATCC	GTAAAAATGT	GTGAGAGTTA	1680
CAAAATGTCT	TCCGGGCTAA	GATCCGACAG	CCATGGTCCA	AAGAAGACTT	CGGCACGTCA	1740
GACTTAAAC	CAGCTTCTA	GCAGAGGCAG	AAGGATCTAG	AGCCAAAGGC	AAAGACTTGA	1800
ATAGGCTGGG	AAGATGCAAG	AATGGCATTT	TACATAAAGA	ACACTCTCTC	CTTTTCCAGC	1860
CAGCACACTT	GCATAGAAAT	TAAGTTTAC	ACTTGAAGTT	CTTGTGTTCC	ATCCTGAGAA	1920
GCTCCAAAGT	CTGAGGTGGT	GTGGTATGCT	GGGTAAATTCT	CCCCACCCCG	CAACATTCCC	1980

TGGGGGTTCC	ATGGGGTAG	CTTCTCCAA	GGACTTCAG	CGGCAACACA	GAAATCCCAC	2040
TTCGAGACAA	AGGAGTTACT	GCTTAAATCA	GGCCCTAATT	TCCAAGGTTC	CCTTGCTTA	2100
AAGTTCCCTA	GAGGACCATC	TCACTTCTAA	AGAAAAGGTG	TATTCGGGGA	CCCATCCTCA	2160
ACCTCCTTGT	TATGGAAGGA	GACTTCGGGA	ACAGAGCAAG	GGCTGAGCCT	CGGGCAGTTT	2220
GGGGTAAGGT	TGGGGTTGGG	GGGAGCAAGG	AAGGCAAGTG	AGGCTGGAGG	CCCAGGGATA	2280
GGGGAAGATG	TGTGTGTGTG	TGTGTGTGTG	TGTGTGTGTG	TGTGTGTGTC	TCGGGGATGG	2340
TGGTGGTGG	CAACTAGGAA	ACTCTGGCGC	TTTCTCCTCC	CCTCACAAAA	CTGAGTCCAG	2400
CTGGAGGCCGC	CTCCAGACTC	TCTGGCCAGG	GCCTCAGAGT	GGTCAACAGT	CCCTGGCCAG	2460
CGTTGCTCTC	TCCAGGCTAA	GGGCACCCAC	TCCCCCTGGAG	ATTCTGAAC	CTGGGCCAGG	2520
AAGAGCCGAA	TTAGACAAGT	GTCTCCAATC	CGGCTGCGTG	CGGATTTGT	TGCGGTGTCC	2580
CTCGGTTGTC	TGCAGTTCT	TTAGTCCCT	CCCTGGCCTG	CCCCTTACAC	CTCCACACAG	2640
GTCCCCCTCT	GTGTAGGAAT	ACACCAAGACC	CTCTCTTAGC	CACACACACC	TCCAGTCCCC	2700
CGTCTACCTA	GATTTTTTC	ATAGCTAGTT	GGATGGGGGA	TGGGTTAGGG	AGGCTGGGTT	2760
TGCGAGCCTC	CAGGTGGGAG	TTCACCGACA	GGTACTCCGC	AAAGGAGCTG	GAAGGCAGGT	2820
CTGGAAAAC	GTCCCCCAGA	TTTAGGATT	TGGGCAGCTT	CCATCAGCTT	ATACTTTGGC	2880
TCCCCCGCCC	CCTAAACTCC	CCATCCCCAC	CTTCCTTCT	CCC GTTACTT	CGTCCTCCCT	2940
CGCCTTTCCA	GCCTTGAGTC	TAAAGCTCCA	TGCTTATGCC	TCTGCAAACA	ACCCCCCTCCC	3000
TTCTAACCCC	AGCAGAACTC	CGAGGAAAGG	GGCCGGAGGC	CCCCCTTCTC	GCCTGTGGTT	3060
AGAGGGGGCA	GTGTGGCAGT	CCCAAGTGGG	GGCGACCGGA	GGCCGTCTCG	GTGCCCGGCC	3120
CGATCAGGCC	ACTGGGCACA	TCGGGGCGG	GAAGCTGGGC	TCACCAAAGG	GGCGACTGGC	3180
CTTGGCAGGT	GTGGGCTCTG	GTCCGGCCTG	GGCAGGCTCC	GGGGCGGGGG	TCTCAGGTTA	3240
CAGCCCCGCG	GGGGGCTGGG	GGCGGGCCCG	CGGTTTGGGC	TGGTTTGCCA	GCCTTGAG	3300
CGACCGGGAG	CATATAACCG	GAGCCTCTGC	TGGGAGAAGA	CGCAGAGCGC	CGCTGGGCTG	3360
CCGGGTCTCC	TGCCTCCTCC	TCCTGCTCCT	AGAGCCTCCT	GCATGAGGGC	GCGGTAGAGA	3420
CCCGGACCCG	CTCCGTGCTC	TGCCGCCTCG	CCGAGCTTCG	CCCGCAAGCT	GGGAATT	3479

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Pro Arg Cys Gly Val Pro Asp Val
1 5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAGCCAAGAT GCGGGGTTGT CGATGTGGGT GAATACAAT

39

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAAAAAAGCCA AGATGCGGGG GTCCTGATGT GGGTGAATAAC

40

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 98 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGTACCACTA GTAAGCTTAG ATCTCATATG GTCGACCCCG GGGATTCCCT GCAGGGATCC
TCTAGAAGTA CTCCATGGGT ATACATCGAT GCGGCCGC

60

98

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2792 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTCGAGTTA CCACTCCCTA TCAGTGATAG AGAAAAGTGA AAGTCGAGTT TACCACTCCC	60
TATCAGTGAT AGAGAAAAGT GAAAGTCGAG TTTACCACTC CCTATCAGTG ATAGAGAAAA	120
GTGAAAGTCG AGTTTACCACTC CCTCTATCAG TGATAGAGAA AAGTGAAGT CGAGTTACC	180
ACTCCCTATC AGTGATAGAG AAAAGTGAAGA GTCGAGTTA CCACCTCCCTA TCAGTGATAG	240
AGAAAAGTGA AAGTCGAGTT TACCACTCCC TATCAGTGAT AGAGAAAAGT GAAAGTCGAG	300
CTCGGTACCC GGGTCGAGTA GGC GTGTACG GTGGGAGGCC TATATAAGCA GAGCTCGTT	360
AGTGAACCGT CAGATCGCCT GGAGACGCCA TCCACGCTGT TTTGACCTCC ATAGAAGACA	420
CCGGGACCGA TCCAGCCTCC GCGGCCCGA ATTAGCTTGA TATCGAATTG GAGCTCGGTA	480
CCCGGGGATC CTCTAGACAA GATGCATCCA GGGGTCTGG CTGCCTTCCT CTTCTTGAGC	540
TGGACTCATT GTCGGGCCT GCCCCTTCCC AGTGGTGGTG ATGAAGATGA TTTGTCTGAG	600
GAAGACCTCC AGTTTGAGA GCGCTACCTG AGATCATACT ACCATCCTAC AAATCTCGCG	660
GGAATCCTGA AGGAGAATGC AGCAAGCTCC ATGACTGAGA GGCTCCGAGA AATGCAGTCT	720
TTCTTCGGCT TAGAGGTGAC TGGCAAACCTT GACGATAACA CCTTAGATGT CATGAAAAG	780
CCAAGATGCG GGGTTGTCGA TGTGGGTGAA TACAATGTTT TCCCTCGAAC TCTTAAATGG	840
TCCAAAATGA ATTTAACCTA CAGAATTGTG AATTACACCC CTGATATGAC TCATTCTGAA	900
GTCGAAAAGG CATTCAAAAA AGCCTTCAAA GTTGGTCCG ATGTAACTCC TCTGAATT	960
ACCAAGACTTC ACGATGGCAT TGCTGACATC ATGATCTCTT TTGGAATTAA GGAGCATGGC	1020
GACTTCTACC CATTGATGG GCCCTCTGGC CTGCTGGCTC ATGCTTTCC TCCTGGGCCA	1080
AATTATGGAG GAGATGCCA TTTTGATGAT GATGAAACCT GGACAAGTAG TTCCAAAGGC	1140
TACAACTTGT TTCTTGTG TGCGCATGAG TTCGGCCACT CCTTAGGTCT TGACCACTCC	1200
AAGGACCTG GAGCACTCAT GTTCCCTATC TACACCTACA CCGGCAAAAG CCACTTTATG	1260
CTTCCTGATG ACGATGTACA AGGGATCCAG TCTCTCTATG GTCCAGGAGA TGAAGACCCC	1320
AACCCTAAAC ATCCAAAAAC GCCAGACAAA TGTGACCCCTT CCTTATCCCT TGATGCCATT	1380
ACCAAGTCTCC GAGGAGAAC AATGATCTT AAAGACAGAT TCTTCTGGCG CCTGCATCCT	1440
CAGCAGGGTTG ATGCGGAGCT GTTTTAACG AAATCATT TT GGCCAGAACT TCCCAACCCT	1500
ATTGATGCTG CATATGAGCA CCCTTCTCAT GACCTCATCT TCATCTTCAG AGGTAGAAA	1560
TTTTGGGCTC TTAATGGTTA TGACATTCTG GAAGGTTATC CCAAAAAAAT ATCTGAAC	1620
GGTCTTCCAA AAGAAGTTAA GAAGATAAGT GCAGCTGTT ACCTTGAGGA TACAGGCAAG	1680
ACTCTCCTGT TCTCAGGAA CCAGGTCTGG AGATATGATG ATACTAACCA TATTATGGAT	1740
AAAGACTATC CGAGACTAAT AGAAGAAGAC TTCCCAGGAA TTGGTATAA AGTAGATGCT	1800
GTCTATGAGA AAAATGGTTA TATCTATT TTCAACGGAC CCATACAGTT TGAATACAGC	1860
ATCTGGAGTA ACCGTATTGT TCGCGTCATG CCAGCAAATT CCATTTGTG GTGTTAAGTG	1920
TCTTTTAAA AATTGTTATT TAAATCCTGA AGAGCATTG GGGTAATACT TCCAGAAGTG	1980
CGGGGTAGGG GAAGAAGAGC TATCAGGAGA AAGCTCTAGT TCTAGAGGGC CCTATTCTAT	2040
AGTGTCACCT AAATGCTAGA GGATCTTGT GAAGGAACCT TACTTCTGTG GTGTGACATA	2100
ATTGGACAAA CTACCTACAG AGATTTAAAG CTCTAAGGTA AATATAAAAT TTTTAAGTGT	2160
ATAATGTGTT AAACACTACTGA TTCTAATTGT TTGTTGATT TAGATTCCAA CCTATGGAAC	2220
TGATGAATGG GAGCAGTGGT GGAATGCCCT TAATGAGGAA AACCTGTTT GCTCAGAAGA	2280
AATGCCATCT AGTGATGATG AGGCTACTGC TGACTCTCAA CATTCTACTC CTCCAAAAAA	2340
GAAGAGAACAG GTAGAAGAGCC CCAAGGACTT TCCTTCAGAA TTGCTAAGTT TTTTGAGTCA	2400
TGCTGTGTT AGTAATAGAA CTCTTGCTTG CTTTGCTATT TACACCACAA AGGAAAAAGC	2460
TGCACTGCTA TACAAGAAAA TTATGGAAA ATATTTGATG TATAGTGCCT TGACTAGAGA	2520
TCATAATCAG CCATACCACA TTTGTAGAGG TTTTACTTG TTTAAAAAAC CTCCCACACC	2580
TCCCCCTGAA CCTGAAACAT AAAATGAATG CAATTGTTGT TGTTAACTTG TTTATTGCAG	2640
CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATT CACAAATAAA GCATTTTTT	2700
CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGGATCA	2760

TCCCGCCATG GGTATAACATC GATGCGGCCG CC

2792

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5276 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGTACCACTA	GTAAGCTTAG	ATCCACTGTC	TGGGATTATA	TCAGGACAAC	CGAACGCCTGG	60
AAAGTGTATT	AGGTAGAGCA	TTTCTTCCA	CGTGTGGGG	CACGTTCCG	ACAGCTAGGA	120
TTCCAGCTCT	GTCTTGAT	GTACAGACT	GTAAATCAAT	CGCAGGTGAA	ACTGTTGGA	180
CAGTAGGTGG	GGATCAAAGA	CCCTCCGCC	GTGAGACTCT	AGGCGCTTTC	CCCTGCCACC	240
AGCCTGTCTC	CAGAGATGCT	CTGGAAGGAG	GCGGGCCCGG	GCGGTCTTTC	TGCTCTTAG	300
CGTGGCGGAC	GGGGCGGCGG	GGCAGGGCT	GGAGCAGAGA	GCGCTGCAGT	GATAGAACTT	360
TCTGACCCCCG	CTGCGCAGGG	CGGCAGGGTG	GCAGGGTGGC	AGGGTGGCGA	GCTAAGCCAG	420
AGCCGAACGC	TGGAGCTCTG	GGAGGAACAT	CGAAGTGT	GTATGTGGTC	TGAGATCGGC	480
CTGACTATAT	TTTTTGTC	TAAATTGCA	AGCACACACC	CACAAAGCTG	CGGTCTTGAC	540
CGGTATTCTT	TATAGAGCGC	AATGGAGTGA	GCTGAGTGT	TAAACGATT	CCCTAATTCA	600
TCTGATAGCA	GAGGCCTCT	CCTAATTGGC	GAAGAGCTGC	CTCATGTCCG	CAACTTTTG	660
GCAGAGTGAA	TTCCACAGCT	TTGTGTGT	GTGTGGGGGG	GGGTGTAAGG	GGTGTCTAAA	720
ACTTTCGGTC	TCCTACTATT	CTGTATCTCG	ACCGGTTGGT	TTTACACCCC	GGCTCATCTC	780
ATCAACGCAA	ACACCCCCAC	TCTCCTATGG	ACCCAAGGAC	CTGACGTGGG	GGAAGGTGGA	840
CATTAGGAAT	GTCAGAAACC	TAGAGTCCAC	GCTCCCTCTC	TCCATCTTTC	CACGAGTTTG	900
GGAAACTTCT	TGGCTGCGAA	GACTTGACCC	CACATCTGCA	TTTCTCAGCC	CCAGCTTCCA	960
AAAGTGTGC	AGGTTGGGA	GGGGAGACCT	CAGTCCTCCT	TTGTGAGGCT	TGTTGCGTT	1020
GAGGGATTGG	CAGCGATGGC	TTCCAGATGG	GCTGAAACCC	TGCCCCTATT	TATTTAAACT	1080
GGTCCTCGT	GGAGAGCTGT	GAATCGGGCT	CTGTATGCGC	TCGAGAAAAG	CCCCATTCTAT	1140
GAGAGGCAAG	GCCCCAGTGGG	TCCCCCCGAC	TCCCCGACCC	CCCTCTCCCA	CAATATATCC	1200
CCCCCTCCCTG	TGCCCGCCTG	CCGCCACCTC	CCGGGCTCCG	GCCCCGCGCG	CAGCGGCGAC	1260
GAAGCAACAC	AGTTCCCCGA	AAGAGGTAGC	TTTTAATTG	GCCAGCCACA	AAGAATCACT	1320
TATGCCGCAC	GGCGGTAACG	AGGGGAACCG	GATCGGGCGG	CCAGGATGCT	ATCTGTGTAG	1380
CCCTTTCTGT	GCCACAATTA	GGGTGGTGCT	GGCTTCCCTC	GACCGCACCT	AGGCGATCTG	1440
GTTACACTGT	TGGCTCCTT	CTTGGGCAGT	CATTTAATCC	TACTTTTAC	TCTACGAATG	1500
TCTGTCTGAT	GGAGGGCTGT	GTCCGGAGCC	CCATCCACAA	AGAGTCAGCC	AGCAGCTCTC	1560
ACACCCGGCT	GGATCTCATA	TGGTGCACTC	TCAGTACAAT	CTGCTCTGAT	GCCGCATAGT	1620
TAAGCCAGCC	AAGCTAGCTT	GCGCAAGCTA	GCTTGCATC	CGTAAAAATG	TGTGAGAGTT	1680
ACAAAATGTC	TTCCGGGCTA	AGATCCGACA	GCCATGGTCC	AAAGAAGACT	TCGGCACTGC	1740
AGACTTAAAA	CCAGCTTCT	AGCAGAGGCA	GAAGGATCTA	GAGCCAAAGG	CAAAGACTTG	1800
AATAGGCTGG	GAAGATGCAA	GAATGGCATT	TTACATAAAAG	AACACTCTCT	CCTTTCCAG	1860
CCAGCACACT	TGCATAGAAA	TTAAGTTTA	CACTTGAAGT	TCTTGTTC	CATCCTGAGA	1920

AGCTCCAAAG	TCTGAGGTGG	TGTGGTATGC	TGGGTAATT	TCCCCACCCC	CCAACATTCC	1980
CTGGGGGTT	CATGGGGTA	GCTTCTCCC	AGGACTTCCA	CGGGCAACAC	AGAAATCCC	2040
CTTCGAGACA	AAGGAGTTAC	TGCTTAAATC	AGGCCCTAAT	TTCCAAGGTT	CCCTTGCTT	2100
AAAGTTCCCT	AGAGGACCAC	CTCACTTCTA	AAGAAAAGGT	GTATTGGGG	ACCCATCCTC	2160
AACCTCCTT	TTATGGAAGG	AGACTTCGGG	AACAGAGCAA	GGGCTGAGCC	TCCGGCAGTT	2220
TGGGGTAAGG	TTGGGGTTGG	GGGGAGCAAG	GAAGGCAAGT	GAGGCTGGAG	GCCCAGGGAT	2280
AGGGGAAGAT	GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT	CTCGGGGATG	2340
GTGGTGGTGG	ACAACTAGGA	AACTCTGGCG	CTTCTCCTC	CCCTCACAAA	ACTGAGTCCA	2400
GCTGGAGCCG	CCTCCAGACT	CTCTGGCCAG	GGCCTCAGAG	TGGTCAACAG	TCCCTGGCCA	2460
GCGTTGCTCT	CTCCAGGCTA	AGGGCACCCA	CTCCCCTGGA	GATTCTGAA	CCTGGGCCAG	2520
GAAGAGCCGA	ATTAGACAAG	TGTCTCCAAT	CCGGCTGCGT	GGGGATTITG	TTGCGGTGTC	2580
CCTCGGTTGT	CTGCAGTTC	TTTAGTCCCT	TCCCTGGCCT	GCCCCTTACA	CCTCCACACAC	2640
GGTCCCCCTC	TGTGTAGGAA	TACACCAGAC	CCTCTCTTAG	CCACACACAC	CTCCAGTCCC	2700
CCGTCTACCT	AGATTTTTT	CATAGCTAGT	TGGATGGGGG	ATGGGTTAGG	GAGGCTGGGT	2760
TTGCGAGCCT	CCAGGTGGGA	GTTCACCGAC	AGGTACTCCG	CAAAGGAGCT	GGAAGGCAGG	2820
TCTGGAAAAC	TGTCCCCCAG	ATTAGGATT	CTGGGCAGCT	TCCATCAGCT	TATACTTTGG	2880
CTCCCCCGCC	CCCTAAACTC	CCCATCCCCA	CCTTCCTTTC	TCCCCTTA	TCGTCTCCC	2940
TCGCCTTCC	AGCCTTGAGT	CTAAAGCTCC	ATGCTTATGC	CTCTGCAAAC	AACCCCTCC	3000
CTTCTAACCC	CAGCAGAACT	CCGAGGAAAG	GGGCCGGAGG	CCCCCCTTCT	CGCCTGTGGT	3060
TAGAGGGGGC	AGTGTGGCAG	TCCCAAGTGG	GGGCGACCGG	AGGCCGTCTC	GGTCCCCCGC	3120
CCGATCAGGC	CACTGGGCAC	ATCGGGGGCG	GGAAGCTGGG	CTCACCAAAG	GGGCGACTGG	3180
CCTTGGCAGG	TGTGGGCTCT	GGTCCGGCCT	GGGCAGGCTC	CGGGGGCGGG	GTCTCAGGTT	3240
ACAGCCCCGC	GGGGGGCTGG	GGGGCGGCC	GCGGTTGGG	CTGGTTTGCC	AGCCTTTGGA	3300
GCGACCGGGA	GCATATAACC	GGAGCCTCTG	CTGGGAGAAG	ACGCAGAGCG	CCGCTGGGCT	3360
GCCGGGTCTC	CTGCCTCCTC	CTCCTGCTCC	TAGAGCCTCC	TGCATGAGGG	CGCGGTAGAG	3420
ACCCGGACCC	GCTCCGTGCT	CTGCCGCCTC	GCCGAGCTTC	GCCCAGCAAGC	TGGGAATT	3480
ATATGTCTAG	ATTAGATAAA	AGTAAAGTGA	TTAACAGCGC	ATTAGAGCTG	CTTAATGAGG	3540
TCGGAATCGA	AGGTTAACCA	ACCCGTAAAC	TCGCCAGAA	GCTAGGTGTA	GAGCAGCCTA	3600
CATTGTATTG	GCATGTAAAA	AATAAGCGGG	CTTGCTCGA	CGCCTTAGCC	ATTGAGATGT	3660
TAGATAGGCA	CCATACTCAC	TTTGCCCTT	TAGAAGGGGA	AAGCTGGCAA	GATTTTTAC	3720
GTAATAACGC	TAAGGTTTT	AGATGTGCTT	TACTAAGTCA	TCGCGATGGA	GCAAAAGTAC	3780
ATTTAGGTAC	ACGGCCTACA	GAAAAAACAGT	ATGAAACTCT	CGAAAATCAA	TTAGCCTTTT	3840
TATGCCAACAA	AGGTTTTCA	CTAGAGAATG	CATTATATGC	ACTCAGCGCT	GTGGGCATT	3900
TTACTTTAGG	TTGCGTATTG	GAAGATCAAG	AGCATCAAGT	CGCTAAAGAA	GAAAGGGAAA	3960
CACCTACTAC	TGATAGTATG	CCGCCATTAT	TACGACAAGC	TATCGAATT	TTTGATCACC	4020
AAGGTGCAGA	GCCAGCCTC	TTATTCGGCC	TTGAATTGAT	CATATCGGGA	TTAGAAAAAC	4080
AACTTAAATG	TGAAAGTGGG	TCCCGTACA	GCCGCGCGCG	TACGAAAAC	AATTACGGGT	4140
CTACCATCGA	GGGCCTGCTC	GATCTCCC	ACGACGACGC	CCCCGAAGAG	GCGGGGCTGG	4200
CGGGCTCCCG	CCTGTCTTT	CTCCCCGCGG	GACACACGCG	CAGACTGTCC	ACGGCCCC	4260
CGACCGATGT	CAGCCTGGGG	GACGAGCTCC	ACTTAGACGG	CGAGGACGTG	GCGATGGCGC	4320
ATGCCGACGC	GCTAGACGGAT	TTCGATCTGG	ACATGTTGGG	GGACGGGGAT	TCCCCGGGTC	4380
CGGGATTTAC	CCCCCACGAC	TCCGCCCCCT	ACGGCGCTCT	GGATATGGCC	GACTTCGAGT	4440
TTGAGCAGAT	GTTCACCGAT	GCCCTTGGAA	TTGACGAGTA	CGGTGGGTAG	GGGGCGCGAG	4500
GATCCTCTAG	AGGGCCCTAT	TCTATAGTGT	CACCTAAATG	CTAGAGGATC	TTTGTGAAGG	4560
AACCTTACTT	CTGTGGTGTG	ACATAATTGG	ACAAACTACC	TACAGAGATT	TAAAGCTCTA	4620
AGGTAAATAT	AAAATTTTA	AGTGTATAAT	GTGTTAAACT	ACTGATTCTA	ATTGTTGTG	4680

TATTTAGAT	TCCAACCTAT	GGAACGTGATG	AATGGGAGCA	GTGGTGGAAAT	GCCTTAATG	4740
AGGAAAACCT	GTTTGCTCA	GAAGAAATGC	CATCTAGTGA	TGATGAGGCT	ACTGCTGACT	4800
CTCAACATTC	TACTCCTCCA	AAAAAGAAGA	GAAAGGTAGA	AGACCCCAAG	GACTTCCCTT	4860
CAGAATTGCT	AAGTTTTTG	AGTCATGCTG	TGTTTAGTAA	TAGAACTCTT	GCTTGCTTTG	4920
CTATTTACAC	CACAAAGGAA	AAAGCTGCAC	TGCTATACAA	GAAAATTATG	GAAAATATT	4980
TGATGTATAG	TGCCTTGACT	AGAGATCATA	ATCAGCCATA	CCACATTTGT	AGAGGTTTA	5040
CTTGCTTTAA	AAAACCTCCC	ACACCTCCCC	CTGAACCTGA	AACATAAAAT	GAATGCAATT	5100
GTTGTTGTTA	ACTTGTATAT	TGAGCTTACA	AATAAAGCAA	TAGCATCACA	5160	
AATTCACAA	ATAAACGATT	TTTTCACTG	CATTCTAGTT	GTGGTTTGTC	CAAACTCATC	5220
AATGTATCTT	ATCATGTCTG	GATCATCCCG	CCATGGGTAT	ACATCGATGC	GGCCGC	5276

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7664 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGTACCACTA	GTAAGCTTAG	ATCCACTGTC	TGGGATTATA	TCAGGACAAC	CGAACCTGG	60
AAAGTGTATT	AGGTAGAGCA	TTTCTTCCA	CGTGTGAGGG	CACGTTCCG	ACAGCTAGGA	120
TTCCAGCTCT	GTCTTGAT	GTACAGACT	GTAAATCAAT	CGCAGGTGAA	ACTGTTGGA	180
CAGTAGGTGG	GGATCAAAGA	CCCTCCGCC	GTGAGACTCT	AGGCCTTTC	CCCTGCCACC	240
AGCCTGTCTC	CAGAGATGCT	CTGGAAGGAG	GCAGGCCCCGG	GCGGTCTTTC	TGCTCTTAG	300
CGTGGCGGAC	GCGGCCGG	GGCAGGGCT	GGAGCAGAGA	GCGCTGCAGT	GATAGAACTT	360
TCTGACCCCG	CTGCGCAGGG	CGGCAGGGTG	GCAGGGTGGC	AGGGTGGCGA	GCTAAGCCAG	420
AGCCGAAACGC	TGGAGCTCTG	GGAGGAACAT	CGAAGTGT	GTATGTGGTC	TGAGATCGGC	480
CTGACTATAT	TTTTTGTC	TAAATTGCA	AGCACACACC	CACAAAGCTG	CGGTCTTGAC	540
CGGTATTCTT	TATAGAGCGC	AATGGAGTGA	GCTGAGTGT	TAAACGATT	CCCTAATTCA	600
TCTGATAGCA	GAGGCCTCT	CCTAATTGGC	GAAGAGCTGC	CTCATGTCCG	CAACTTTTG	660
GCAGAGTGAA	TTCCACAGCT	TTGTGTGT	GTGTGGGGGG	GGGTGTAAGG	GGTGTCTAAA	720
ACTTTCGGTC	TCCTACTATT	CTGTATCTCG	ACCGGTTGGT	TTTACACCCC	GGCTCATCTC	780
ATCAACGCAA	ACACCCCCAC	TCTCCTATGG	ACCCAAGGAC	CTGACGTGGG	GGAAGGTGGA	840
CATTAGGAAT	GTCAGAAACC	TAGAGTCCAC	GCTCCTCCTC	TCCATCTTTC	CACGAGTTG	900
GGAAACTTCT	TGGCTGCGAA	GACTTGACC	CACATCTGCA	TTTCTCAGCC	CCAGCTTCCA	960
AAAGTGTGTC	AGGTTCGGGA	GGGGAGACCT	CAGTCCTCCT	TTGTGAGGCT	TGTTTGCCTT	1020
GAGGGATTGG	CAGCGATGGC	TTCCAGATGG	GCTGAAACCC	TGCCCGTATT	TATTTAAACT	1080
GGTTCCTCGT	GGAGAGCTGT	GAATCGGGCT	CTGTATGCGC	TCGAGAAAAG	CCCCATTAT	1140
GAGAGGCAAG	GCCCCAGTGGG	TCCCCCCGAC	TCCCCGACCC	CCCTCTCCCA	CAATATATCC	1200
CCCCTCCCTG	TGCCCGCTG	CCGCCACCTC	CCGGGCTCCG	GCCCCGCGCG	CAGCGGCGAC	1260
GAAGCAACAC	AGTTCCCCGA	AAGAGGTAGC	TTTTAATTG	GCCAGCCACA	AAGAATCACT	1320
TATGCCGCAC	GGCGGTAACG	AGGGGAACCG	GATCGGGCGG	CCAGGATGCT	ATCTGTGTAG	1380

CCCTTTCGT	GCCACAATTA	GGGTGGTGCT	GGCTTCCTCC	GACCGCACCT	AGGCGATCTG	1440
GTTACACTGT	TGGCTCCTT	CTTGGGCAGT	CATTTAATCC	TACTTTTAC	TCTACGAATG	1500
TCTGTCTGAT	GGAGGGCTGT	GTCGGAGCC	CCATCCACAA	AGAGTCAGCC	AGCAGCTCTC	1560
ACACCCGGCT	GGATCTCATA	TGGTGCACTC	TCAGTACAAT	CTGCTCTGAT	GCCGCATAGT	1620
TAAGCCAGCC	AAGCTAGCTT	GCGCAAGCTA	GCTTGCATC	CGTAAAATG	TGTGAGAGTT	1680
ACAAAATGTC	TTCCGGGCTA	AGATCCGACA	GCCATGGTCC	AAAGAAGACT	TCGGCACTGC	1740
AGACTTAAAA	CCAGCTTCT	AGCAGAGGCA	GAAGGATCTA	GAGCCAAAGG	CAAAGACTTG	1800
AATAGGCTGG	GAAGATGCAA	GAATGGCATT	TTACATAAAG	AACACTCTCT	CCTTTCCAG	1860
CCAGCACACT	TGCATAGAAA	TTAAGTTTA	CACTTGAAGT	TCTTTGTTTC	CATCCTGAGA	1920
AGCTCCAAAG	TCTGAGGTGG	TGTGGTATGC	TGGGTAATT	TCCCCACCCC	CCAACATTCC	1980
CTGGGGGTTC	CATGGGGTA	GCTTCTCCA	AGGACTTCCA	GCGGCAACAC	AGAAATCCCA	2040
CTTCGAGACA	AAGGAGTTAC	TGCTTAAATC	AGGCCCTAAT	TTCCAAGGTT	CCCTTGCTT	2100
AAAGTTCCCT	AGAGGACCAT	CTCACTTCTA	AAGAAAAGGT	GTATTGGGG	ACCCATCCTC	2160
AACCTCCTTG	TTATGGAAGG	AGACTTCGGG	AACAGAGCAA	GGGCTGAGCC	TCCGGCAGTT	2220
TGGGGTAAGG	TTGGGGTTGG	GGGGAGCAAG	GAAGGCAAGT	GAGGCTGGAG	GCCCAGGGAT	2280
AGGGGAAGAT	GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT	CTCGGGGATG	2340
GTGGTGGTGG	ACAACCTAGGA	AACTCTGGCG	CTTCTCCTC	CCCTCACAAA	ACTGAGTCCA	2400
GCTGGAGCCG	CCTCCAGACT	CTCTGGCCAG	GGCCTCAGAG	TGGTCAACAG	TCCCTGGCCA	2460
GCGTTGCTCT	CTCCAGGCTA	AGGGCACCCA	CTCCCCTGGA	GATTCTGAA	CCTGGGCCAG	2520
GAAGAGCCGA	ATTAGACAAG	TGTCTCCAAT	CCGGCTGCGT	GC GGATTTG	TTGCGGTGTC	2580
CCTCGGTTGT	CTGCAGTTCC	TTTAGTCCCT	TCCCTGGCCT	GCCCCTTACA	CCTCCACACA	2640
GGTCCCCCTC	TGTGTAGGAA	TACACCAGAC	CCTCTCTTAG	CCACACACAC	CTCCAGTCCC	2700
CCGTCTACCT	AGATTTTTT	CATAGCTAGT	TGGATGGGGG	ATGGGTTAGG	GAGGCTGGGT	2760
TTGCAGCCT	CCAGGTGGGA	GTTCACCGAC	AGGTACTCCG	CAAAGGAGCT	GGAAGGCAGG	2820
TCTGGAAAAC	TGTCCCCCAG	ATTAGGATT	CTGGGCAGCT	TCCATCAGCT	TATACTTTGG	2880
CTTCCCCGCC	CCCTAAACTC	CCCATCCCCA	CCTTCCTTTC	TCCCCTACT	TCGTCTCCCC	2940
TCGCCTTCC	AGCCTTGAGT	CTAAAGCTCC	ATGCTTATGC	CTCTGCAAAC	AACCCCTCC	3000
CTTCTAACCC	CAGCAGAACT	CCGAGGAAAG	GGGCCGGAGG	CCCCCCTCT	CGCCTGTGGT	3060
TAGAGGGGGC	AGTGTGGCAG	TCCCAAGTGG	GGGCGACCGG	AGGCCGTCTC	GGTCCCCCGC	3120
CCGATCAGGC	CACTGGCAC	ATCGGGGGCG	GGAAGCTGGG	CTCACCAAAG	GGGCGACTGG	3180
CCTTGGCAGG	TGTGGGCTCT	GGTCCGGCCT	GGGCAGGCTC	CGGGGGCGGG	GTCTCAGGTT	3240
ACAGCCCCGC	GGGGGGCTGG	GGGGCGGCC	GCGGTTGGG	CTGGTTGCC	AGCCTTGGGA	3300
GCGACCGGGA	GCATATAACC	GGAGCCTCTG	CTGGGAGAAG	ACGCAGAGCG	CCGCTGGGCT	3360
GCCGGGTCTC	CTGCCTCCTC	CTCCTGCTCC	TAGAGCCTCC	TGCATGAGGG	CGCGGTAGAG	3420
ACCCGGACCC	GCTCCGTGCT	CTGCCGCCTC	GCCGAGCTTC	GCCCGCAAGC	TGGGAATT	3480
GGATCCCCGG	GATCGAAAGA	GCCTGCTAA	GCAAAAAAGA	AGTCACCATG	TCGTTACTT	3540
TGACCAACAA	GAACGTGATT	TTCGTTGCCG	GTCTGGGAGG	CATTGGTCTG	GACACCAGCA	3600
AGGAGCTGCT	CAAGCGCGAT	CCCGTCGTT	TACAACGTG	TGACTGGAA	AACCTGGCG	3660
TTACCCAAC	TAATCGCCTT	GCAGCACATC	CCCCTTCGC	CAGCTGGCTT	TATAGCGAAG	3720
AGGCCCGCAC	CGATGCCCT	TCCCAACAGT	TGCGCAGCCT	GAATGGCGAA	TGGCGCTTTG	3780
CCTGGTTTCC	GGCACCAAGAA	GGGGTGCCTGG	AAAGCTGGCT	GGAGTGCAGT	CTTCCTGAGG	3840
CCGATACTGT	CGTCGTCCCC	TCAAACTGGC	AGATGCACGG	TTACGATGCG	CCCATCTACA	3900
CCAACGTAAC	CTATTCCATT	ACGGTCAATC	CGCCGTTTGT	TCCCACGGAG	AATCCGACGG	3960
GTTGTTACTC	GCTCACATT	AATGTTGATG	AAAGCTGGCT	ACAGGAAGGC	CAGACGCGAA	4020
TTATTTTGA	TGGCGTTAAC	TTGGCGTTTC	ATCTGTGGTG	CAACGTGCGC	TGGGTCGGTT	4080
ACGGCCAGGA	CAGTCGTTTG	CCGTCTGAAT	TTGACCTGAG	CGCATTTTA	CGCGCCGGAG	4140

AAAACCGCCT	CGCGGTGATG	GTGCTGCCTT	GGAGTGACGG	CAGTTATCTG	GAAGATCAGG	4200
ATATGTGGCG	GATGAGCGGC	ATTTCCTCGT	ACGTCTCGTT	GCTGCATAAA	CCGACTACAC	4260
AAATCAGCGA	TTTCCATGTT	GCCACTCGCT	TTAATGATGA	TTTCAGCCGC	GCTGAACTGG	4320
AGGCTGAAGT	TCAGATGTGC	GGCGAGTTGC	GTGACTACCT	ACGGGTAACA	GTTTCTTTAT	4380
GGCAGGGTGA	AACGCAGGTC	GCCAGCGGC	CCGCGCCTT	CGGCGGTGAA	ATTATCGATG	4440
AGCGTGGTGG	TTATGCCGAT	CGCGTCACAC	TACGTCTGAA	CGTCGAAAAC	CCGAAACTGT	4500
GGAGCGCCGA	AATCCCAGAAT	CTCTATCGTG	CGGTGGTTGA	ACTGCACACC	GCCGACGGCA	4560
CGCTGATTGA	AGCAGAACGC	TGCGATGTG	GTTCCGCGA	GGTGGGATT	GAAAATGGTC	4620
TGCTGCTGCT	GAACGGCAAG	CCGTTGCTGA	TTCGAGGC	TAACCGTCAC	GAGCATCATC	4680
CTCTGCATGG	TCAGGTCA	GATGAGCAGA	CGATGGTGCA	GGATATCCTG	CTGATGAAGC	4740
AGAACAACTT	TAACGCCGT	CGCTGTTCGC	ATTATCCGA	CCATCCGCTG	TGGTACACGC	4800
TGTGCGACCG	CTACGGCCTG	TATGTGGTGG	ATGAAGCCAA	TATTGAAACC	CACGGCATGG	4860
TGCCAATGAA	TCTGCTGACC	GATGATCCGC	GCTGGCTACC	GGCGATGAGC	GAACCGTAA	4920
CGCGAATGGT	GCAGCGCGAT	CGTAATCACC	CGAGTGTGAT	CATCTGGTCG	CTGGGAAATG	4980
AATCAGGCCA	CGGCGCTAAT	CACGACGCGC	TGTATCGCTG	GATCAAATCT	GTCGATCCTT	5040
CCCGCCCCGT	GCAGTATGAA	GGCGGCGGAG	CCGACACCAC	GGCCACCGAT	ATTATTTGCC	5100
CGATGTACGC	GGCGGTGGAT	GAAGACCAGC	CCTTCCCGGC	TGTGCCGAAA	TGGTCCATCA	5160
AAAAATGGCT	TTCGCTACCT	GGAGAGACGC	GCCCGCTGAT	CCTTTGCGAA	TACGCCACG	5220
CGATGGGTA	CAGTCTGGC	GGTTTCGCTA	AATACTGGCA	GGCGTTTCGT	CAGTATCCCC	5280
GTTTACAGGG	CGGCTTCGTC	TGGGACTGGG	TGGATCAGTC	GCTGATTAAA	TATGATGAAA	5340
ACGGCAACCC	GTGGTCGGCT	TACGGCGGTG	ATTTCGGCGA	TACGCCAAC	CATGCCAGT	5400
TCTGTATGAA	CGGTCTGGC	TTTGGCGACC	GCACGCCGCA	TCCAGCGCTG	ACGGAAGCAA	5460
AACACCAGCA	GCAGTTTTTC	CAGTTCCGTT	TATCCGGGCA	AACCATCGAA	GTGACCAGCG	5520
AATAACCTGTT	CCGTCTAGC	GATAACGAGC	TCCTGCACTG	GATGGTGGCG	CTGGATGGTA	5580
AGCCGCTGGC	AAGCGGTGAA	GTGCCTCTGG	ATGTCGCTCC	ACAAGGTAAA	CAGTTGATTG	5640
AACTGCCTGA	ACTACCGCAG	CCGGAGAGCG	CCGGGCAACT	CTGGCTCACA	GTACCGTAG	5700
TGCAACCGAA	CGCGACCGGA	TGGTCAGAAG	CCGGGCACAT	CAGCGCTGG	CAGCAGTGGC	5760
GTCTGGCGGA	AAACCTCAGT	GTGACGCTCC	CCGCCGCGTC	CCACGCCATC	CCGCATCTGA	5820
CCACCAGCGA	AATGGATTAA	TGCATCGAGC	TGGGTAATAA	GCGTTGGCAA	TTAACCGCC	5880
AGTCAGGCTT	TCTTCACAG	CTGTGGATTG	GCGATAAAAAA	ACAACGTCTG	ACGCCGCTGC	5940
GCGATCAGTT	CACCCGTGCA	CCGCTGGATA	ACGACATTGG	CGTAAGTGAA	GCGACCCGCA	6000
TTGACCTAA	CGCCTGGGTC	GAACGCTGGA	AGGCAGCGGG	CCATTACCAAG	GCCGAAGCAG	6060
CGTTGTTGCA	GTGCACGGCA	GATACACTG	CTGATCGGGT	GCTGATTACG	ACCGCTCACG	6120
CGTGGCAGCA	TCAGGGGAAA	ACCTTATTAA	TCAGCCGGAA	AACCTACCGG	ATTGATGGTA	6180
GTGGTCAAAT	GGCGATTACC	GTTGATGTTG	AAGTGGCGAG	CGATACACCG	CATCCGGCGC	6240
GGATTGGCCT	GAACGTCCAG	CTGGCGCAGG	TAGCAGAGCG	GGTAAACTGG	CTCGGATTAG	6300
GGCCGCAAGA	AAACTATCCC	GACCGCCTTA	CTGCCGCTG	TTTGACCGC	TGGGATCTGC	6360
CATTGTCAGA	CATGTATACC	CCGTACGCT	TCCCGAGCGA	AAACGGTCTG	CGCTGCGGG	6420
CGCGCGAATT	GAATTATGGC	CCACACCAGT	GGCGCGGCCG	CTTCCAGTTC	AACATCAGCC	6480
GCTACAGTCA	ACAGCAACTG	ATGGAAACCA	GCCATCGCCA	TCTGCTGCAC	GCGGAAGAAG	6540
GCACATGGCT	GAATATCGAC	GGTTTCATA	TGGGGATTGG	TGGCGACGAC	TCCTGGAGCC	6600
CGTCAGTATC	GGCGGAATT	CAGCTGAGCG	CCGGTCGCTA	CCATTACCAAG	TTGGTCTGGT	6660
GTCAAAAATA	ATAATAACCG	GCAGGCCATG	TCTGAAAGTA	TTCGCGTAAG	GAAATCCATT	6720
ATGTACTATT	AAAAAACAC	AAACTTTGG	ATGTTCGGTT	TATTCTTTT	CTTTTACTTT	6780
TTTATCATGG	GAGCCTACTT	CCCGTTTTTC	CCGATTTGGC	TACATGACAT	CAACCATATG	6840
AGCAAAAGTG	ATACGGGTAT	TATTTTGCC	GCTATTCTC	TGTTGTCGCT	ATTATTCCAA	6900

CCGCTGTTGG TCTGCTTCT GACAAACTCG GCCTCGACTC TAGACTGAGA ACTTCAGGGT	6960
GAGTTTGGGG ACCCTTGATT GTTCTTCCTT TTTCGCTATT GAAAAATTCA TGTTATATGG	7020
AGGGGGCAAA GTTTCAAGGG TGTTGTTAG AATGGGAAGA TGTCCTGT ATCACCATGG	7080
ACCCTCATGA TAATTTGTT TCTTCACCT TCTACTCTGT TGACAACCAT TGTCTCCTCT	7140
TATTTCTT TCATTTCTG TAACTTTTT CGTTAAACTT TAGCTTGCAT TTGTAACGAA	7200
TTTTAAATT CACTTCGTT TATTTGTCAG ATTGTAAGTA CTTTCTCTAA TCACCTTTTT	7260
TTCAAGGCAA TCAGGGTAAT TATATTGTAC TTCAGCACAG TTTAGAGAA CAATTGTTAT	7320
AATTAAATGA TAAGGTAGAA TATTTCTGCA TATAAATTCT GGCTGGCGTG GAAATATTCT	7380
TATTGGTAGA AACAACTACA TCCTGGTAAT CATCCTGCCT TTCTCTTTAT GGTTACAATG	7440
ATATACACTG TTTGAGATGA GGATAAAAATA CTCTGAGTCC AAACCGGGCC CCTCTGCTAA	7500
CCATGTTCAT GCCTTCTTCT TTTTCCTACA GCTCCTGGC AACGTGCTGG TTGTTGTGCT	7560
GTCTCATCAT TTTGGCAAAG AATTCACTCC TCAGGTGCAG GCTGCCTATC AGAAGGTGGT	7620
GGCTGGTGTG GCCAATGCC C TGGCTCACAA ATACCACTGA GATC	7664

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGAGGGCCTG CTCGATCTCC

20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGCATTCCAC CACTGCTCCC

20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGCACCCCTT CTCATGACCT C

21

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTTGGTGTAG ATGGGCGCAT CG

22

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCGGGGTCTC AGGTTACAGC C

21

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GCCCTCTGGC CTGCTGGCTC ATG

23

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CAGGAGAGTC TTGCCTGTAT CCTC

24

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1521 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CAAGATGCAT CCAGGGTCC TGGCTGCCTT CCTCTTCTTG AGCTGGACTC ATTGTCGGGC	60
CCTGCCCTT CCCAGTGGTG GTGATGAAGA TGATTTGTCT GAGGAAGACC TCCAGTTG	120
AGAGCGCTAC CTGAGATCAT ACTACCATCC TACAAATCTC GCAGGAATCC TGAAGGAGAA	180
TGCAGCAAGC TCCATGACTG AGAGGCTCG AGAAATGCAG TCTTCTTCG GCTTAGAGGT	240
GACTGGAAA CTTGACGATA ACACCTTAGA TGTATGAAA AAGCCAAGAT GCGGGGTTGT	300
CGATGTGGG TGAATACAATG TTTCCCTCG AACTCTAAA TGGTCCAAAA TGAATTAAAC	360
CTACAGAATT GTGAATTACA CCCCTGATAT GACTCATTCT GAAGTCGAAA AGGCATTCAA	420
AAAAGCCTTC AAAGTTGGT CCGATGTAAC TCCTCTGAAT TTTACCAAGAC TTCACGATGG	480
CATTGCTGAC ATCATGATCT CTTTGGAAT TAAGGAGCAT GGCGACTTCT ACCCATTG	540
TGGGCCCTCT GGCCTGCTGG CTCATGCTTT TCCTCCTGGG CCAAATTATG GAGGAGATGC	600
CCATTTGAT GATGATGAAA CCTGGACAAG TAGTTCCAAA GGCTACAAC TGTGTTCTTGT	660
TGCTGCGCAT GAGTCGGCC ACTCCTTAGG TCTTGACCAC TCCAAGGACC CTGGAGCACT	720
CATGTTTCCT ATCTACACCT ACACCGGCAA AAGCCACTT ATGCTTCTG ATGACGATGT	780
ACAAGGGATC CAGTCTCTCT ATGGTCCAGG AGATGAAGAC CCCAACCTA AACATCCAAA	840
AACGCCAGAC AAATGTGACC CTTCCCTTATC CCTTGATGCC ATTACCAAGTC TCCGAGGAGA	900
AACAATGATC TTTAAAGACA GATTCTTCTG GCGCCTGCAT CCTCAGCAGG TTGATGCGGA	960
GCTGTTTTA ACGAAATCAT TTTGGCCAGA ACTTCCCAAC CGTATTGATG CTGCATATGA	1020

GCACCCTTCT	CATGACCTCA	TCTTCATCTT	CAGAGGTAGA	AAATTTGGG	CTCTTAATGG	1080
TTATGACATT	CTGGAAGGTT	ATCCCAAAAA	AATATCTGAA	CTGGGTCTTC	CAAAAGAAGT	1140
TAAGAAGATA	AGTGCAGCTG	TTCACTTGAA	GGATACAGGC	AAGACTCTCC	TGTTCTCAGG	1200
AAACCAGGTC	TGGAGATATG	ATGATACTAA	CCATATTATG	GATAAAGACT	ATCCGAGACT	1260
AATAGAAGAA	GACTTCCCAG	GAATTGGTGA	TAAAGTAGAT	GCTGTCTATG	AGAAAAATGG	1320
TTATATCTAT	TTTTCAACG	GACCCATACA	GTTGAATAC	AGCATCTGGA	GTAACCGTAT	1380
TGTCGCGTC	ATGCCAGCAA	ATTCCATTAA	GTGGTGTAA	GTGTCTTTT	AAAAATTGTT	1440
ATTTAAATCC	TGAAGAGCAT	TTGGGTAAT	ACTTCCAGAA	GTGCGGGTA	GGGAAAGAAG	1500
AGCTATCAGG	AGAAAGCTTG	G				1521

(2) INFORMATION FOR SEQ ID NO:19

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Pro Arg Cys Gly Xaa Pro Asp

1 5

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

His Glu Xaa Gly His Xaa Xaa Xaa Xaa His Ser

1 5 10

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 471 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met His Pro Gly Val Leu Ala Ala Phe Leu Phe Leu Ser Trp Thr His
1 5 10 15
Cys Arg Ala Leu Pro Leu Pro Ser Gly Gly Asp Glu Asp Asp Leu Ser
20 25 30
Glu Glu Asp Leu Gln Phe Ala Glu Arg Tyr Leu Arg Ser Tyr Tyr His
35 40 45
Pro Thr Asn Leu Ala Gly Ile Leu Lys Glu Asn Ala Ala Ser Ser Met
50 55 60
Thr Glu Arg Leu Arg Glu Met Gln Ser Phe Phe Gly Leu Glu Val Thr
65 70 75 80
Gly Lys Leu Asp Asp Asn Thr Leu Asp Val Met Lys Lys Pro Arg Cys
85 90 95
Gly Gly Val Asp Val Gly Glu Tyr Asn Val Phe Pro Arg Thr Leu Lys
100 105 110
Trp Ser Lys Met Asn Leu Thr Tyr Arg Ile Val Asn Tyr Thr Pro Asp
115 120 125
Met Thr His Ser Glu Val Glu Lys Ala Phe Lys Lys Ala Phe Lys Val
130 135 140
Trp Ser Asp Val Thr Pro Leu Asn Phe Thr Arg Leu His Asp Gly Ile
145 150 155 160
Ala Asp Ile Met Ile Ser Phe Gly Ile Lys Glu His Gly Asp Phe Tyr
165 170 175
Pro Phe Asp Gly Pro Ser Gly Leu Leu Ala His Ala Phe Pro Pro Gly
180 185 190
Pro Asn Tyr Gly Gly Asp Ala His Phe Asp Asp Asp Glu Thr Trp Thr
195 200 205
Ser Ser Ser Lys Gly Tyr Asn Leu Phe Leu Val Ala Ala His Glu Phe
210 215 220
Gly His Ser Leu Gly Leu Asp His Ser Lys Asp Pro Gly Ala Leu Met
225 230 235 240
Phe Pro Ile Tyr Thr Tyr Thr Gly Lys Ser His Phe Met Leu Pro Asp
245 250 255
Asp Asp Val Gln Gly Ile Gln Ser Leu Tyr Gly Pro Gly Asp Glu Asp
260 265 270
Pro Asn Pro Lys His Pro Lys Thr Pro Asp Lys Cys Asp Pro Ser Leu
275 280 285
Ser Leu Asp Ala Ile Thr Ser Leu Arg Gly Glu Thr Met Ile Phe Lys
290 295 300
Asp Arg Phe Phe Trp Arg Leu His Pro Gln Gln Val Asp Ala Glu Leu

305	310	315	320												
Phe	Leu	Thr	Lys	Ser	Phe	Trp	Pro	Glu	Leu	Pro	Asn	Arg	Ile	Asp	Ala
325	330	335													
Ala	Tyr	Glu	His	Pro	Ser	His	Asp	Leu	Ile	Phe	Ile	Phe	Arg	Gly	Arg
340	345	350													
Lys	Phe	Trp	Ala	Leu	Asn	Gly	Tyr	Asp	Ile	Leu	Glu	Gly	Tyr	Pro	Lys
355	360	365													
Lys	Ile	Ser	Glu	Leu	Gly	Leu	Pro	Lys	Glu	Val	Lys	Lys	Ile	Ser	Ala
370	375	380													
Ala	Val	His	Phe	Glu	Asp	Thr	Gly	Lys	Thr	Leu	Leu	Phe	Ser	Gly	Asn
385	390	395	400												
Gln	Val	Trp	Arg	Tyr	Asp	Asp	Thr	Asn	His	Ile	Met	Asp	Lys	Asp	Tyr
405	410	415													
Pro	Arg	Leu	Ile	Glu	Glu	Asp	Phe	Pro	Gly	Ile	Gly	Asp	Lys	Val	Asp
420	425	430													
Ala	Val	Tyr	Glu	Lys	Asn	Gly	Tyr	Ile	Tyr	Phe	Phe	Asn	Gly	Pro	Ile
435	440	445													
Gln	Phe	Glu	Tyr	Ser	Ile	Trp	Ser	Asn	Arg	Ile	Val	Arg	Val	Met	Pro
450	455	460													
Ala	Asn	Ser	Ile	Leu	Trp	Cys									
465	470														